

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5528538号
(P5528538)

(45) 発行日 平成26年6月25日(2014.6.25)

(24) 登録日 平成26年4月25日(2014.4.25)

(51) Int. Cl.

F I

G 1 O L 21/0232 (2013.01)

G 1 O L 21/0232

G 1 O L 21/0264 (2013.01)

G 1 O L 21/0264

Z

請求項の数 3 (全 14 頁)

(21) 出願番号 特願2012-504136 (P2012-504136)
 (86) (22) 出願日 平成22年3月9日(2010.3.9)
 (86) 国際出願番号 PCT/JP2010/001649
 (87) 国際公開番号 W02011/111091
 (87) 国際公開日 平成23年9月15日(2011.9.15)
 審査請求日 平成24年6月12日(2012.6.12)

(73) 特許権者 000006013
 三菱電機株式会社
 東京都千代田区丸の内二丁目7番3号
 (74) 代理人 100123434
 弁理士 田澤 英昭
 (74) 代理人 100101133
 弁理士 濱田 初音
 (74) 代理人 100173934
 弁理士 久米 輝代
 (74) 代理人 100156351
 弁理士 河村 秀央
 (72) 発明者 古田 訓
 東京都千代田区丸の内二丁目7番3号 三
 菱電機株式会社内

最終頁に続く

(54) 【発明の名称】 雑音抑圧装置

(57) 【特許請求の範囲】

【請求項1】

時間領域の入力信号を周波数領域の信号であるパワースペクトルと位相スペクトルに変換する時間・周波数変換部と、

前記入力信号に重畳した雑音スペクトルを推定する雑音スペクトル推定部と、

前記パワースペクトルと前記雑音スペクトルとを用いて、雑音抑圧量を算出する雑音抑圧量生成部と、

前記雑音抑圧量に応じて前記パワースペクトルの振幅を抑圧する雑音抑圧部と、

前記位相スペクトルと前記雑音抑圧部で振幅抑圧された前記パワースペクトルとを時間領域の信号に変換する周波数・時間変換部とを備える雑音抑圧装置において、

前記時間・周波数変換部が変換した複数のパワースペクトルを1グループにまとめ、当該グループ内の前記複数のパワースペクトルのうち、値が大きいものを優先して選択して代表パワースペクトルにする代表成分生成部を備え、

前記雑音抑圧量生成部は、前記代表パワースペクトルを用いて雑音抑圧量を算出し、

前記入力信号が音声らしいかどうかの度合いを示す音声らしさ評価値を算出する音声らしさ推定部を備え、

前記代表成分生成部は、前記音声らしさ評価値に基づいた代表パワースペクトルを生成することを特徴とする雑音抑圧装置。

【請求項2】

代表成分生成部は、音声らしさ評価値に基づいて、入力信号の音声らしさの度合いが高

い場合にはグループ内の値の大きいパワースペクトルを優先して選択して代表パワースペクトルを生成し、当該入力信号の音声らしさの度合いが低い場合には当該グループ内の複数のパワースペクトルの平均値を求めて代表パワースペクトルを生成することを特徴とする請求項1記載の雑音抑圧装置。

【請求項3】

代表パワースペクトルは、音声らしさ評価値を重み付け係数に用いた、グループ内の複数のパワースペクトルのうちの最大値とグループ内の複数のパワースペクトルの平均値の重み付き和であることを特徴とする請求項1記載の雑音抑圧装置。

【発明の詳細な説明】

【技術分野】

10

【0001】

この発明は、音声信号に重畳した雑音を抑圧する雑音抑圧装置に関する。

【背景技術】

【0002】

雑音抑圧装置は、主として、音声信号に雑音が重畳した時間領域の信号を入力信号として入力し、この入力信号を周波数領域の信号であるパワースペクトルに変換した後、入力信号のパワースペクトルから雑音の平均的なパワースペクトルを推定し、推定した雑音のパワースペクトルを入力信号のパワースペクトルから減算して雑音抑圧した入力信号のパワースペクトルを得て、それを元の時間領域の信号に戻すことにより雑音抑圧処理を行う。

20

【0003】

このような従来雑音抑圧装置として、例えば特許文献1が開示されている。特許文献1に開示されている雑音抑圧装置は、非特許文献1に開示されている技術を基本とし、雑音スペクトル推定と抑圧量の算出時に入力信号の複数のパワースペクトル成分の平均値を求め、得られた1つの平均値から雑音スペクトル推定と抑圧量算出を行い、それらを複数のパワースペクトル成分に共通して適用していた。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特許4172530号公報（第8頁～12頁、図2）

30

【非特許文献】

【0005】

【非特許文献1】Y. Ephraim, D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. ASSP, Vol. 32, No. 6, pp. 1109-1121, Dec. 1984

【発明の概要】

【0006】

従来雑音抑圧装置は以上のように構成されているので、以下に述べる課題があった。

40

【0007】

従来雑音抑圧装置では、雑音抑圧のための抑圧量算出において、入力信号のパワースペクトル成分毎にベッセル関数など複雑な計算を行う必要があり、非常に処理量が掛かる。そのため特許文献1に開示された従来雑音抑圧装置では、複数のスペクトル成分をまとめて平均化し、平均化したスペクトル成分を各スペクトル成分の代表スペクトル成分として計算を行うことで処理量の削減を行っている。しかしながら、この方法ではスペクトル成分に振幅が大きい成分（即ち、音声成分と考えられる）が存在しても、平均化することで音声成分が過小に取り扱われることとなり、その結果、音声信号が抑圧されて音声の隠滅感が増大し、音質劣化する課題がある。

【0008】

50

この発明は、かかる課題を解決するためになされたもので、少ない処理量で高品質な雑音抑圧を行うことが可能な雑音抑圧装置を提供することを目的とする。

【0009】

この発明の雑音抑圧装置は、時間・周波数変換部が変換した複数のパワースペクトルを1グループにまとめ、当該グループ内の複数のパワースペクトルのうち、値が大きいものを優先して選択して代表パワースペクトルにする代表成分生成部を備え、雑音抑圧量生成部が代表パワースペクトルを用いて雑音抑圧量を算出し、入力信号が音声らしいかどうかの度合いを示す音声らしさ評価値を算出する音声らしさ推定部を備え、代表成分生成部は、音声らしさ評価値に基づいた代表パワースペクトルを生成するようにしたものである。

【0010】

この発明によれば、代表パワースペクトルを用いて雑音抑圧量を算出するので処理量が少なく済み、かつ、この代表パワースペクトルにはグループ内の値の大きいパワースペクトルを用いるので雑音抑圧量算出時に入力信号の音声成分が過小評価されることがなくなり、その結果、音声信号を抑圧せず、高品質な雑音抑圧を行うことができる。

【図面の簡単な説明】

【0011】

【図1】この発明の実施の形態1に係る雑音抑圧装置の構成を示すブロック図である。

【図2】帯域分離部によるパワースペクトルの帯域分割の一例を示すグラフである。

【図3】帯域代表成分生成部の処理効果を模式的に表し、図3(a)は入力信号のパワースペクトルのグラフ、図3(b)はサブバンド内のパワースペクトルの平均値を代表にする場合(従来法)、図3(c)はサブバンド内のパワースペクトルの最大値を代表にする場合(本発明)である。

【図4】雑音抑圧量生成部の詳細な構成を示すブロック図である。

【発明を実施するための形態】

【0012】

以下、この発明をより詳細に説明するために、この発明を実施するための形態について、添付の図面に従って説明する。

実施の形態1

図1に示す雑音抑圧装置は、入力端子1と、時間・周波数変換部2と、音声らしさ推定部3と、雑音スペクトル推定部4と、帯域分離部5と、帯域代表成分生成部(代表成分生成部)6と、雑音抑圧量生成部7と、帯域多重化部8と、雑音抑圧部9と、周波数・時間変換部10と、出力端子11とを備える。

【0013】

この雑音抑圧装置の入力としては、マイクロホン(図示せず)等を通じて取り込まれた音声及び音楽等がA/D(アナログ・デジタル)変換された後、所定のサンプリング周波数(例えば、8kHz)でサンプリングされると共にフレーム単位(例えば、10ms)に分割された信号を用いる。

【0014】

以下、図1に基づいて、実施の形態1に係る雑音抑圧装置の動作原理を説明する。

入力端子1は、上述のような信号を受け付けて、入力信号 $y(t)$ として時間・周波数変換部2へ出力する。

【0015】

時間・周波数変換部2は、フレーム単位に分割された入力信号 $y(t)$ に対して窓掛け処理を行い、その窓掛け後の信号 $y(n, t)$ に対して、例えば256点のFFT(Fast Fourier Transform: 高速フーリエ変換)を用いて時間軸上の信号を周波数上の信号(スペクトル)に変換し、入力信号のパワースペクトル $Y(n, k)$ と位相スペクトル $P(n, k)$ を算出する。ここで、 n はフレーム番号、 k はスペクトル番号、 t は離散時間番号を表す。以降、特に示す必要が無い限り、現フレームの入力信号を指すものとし、その信号がスペクトルを表す場合にはフレーム番号を省略する。

【0016】

10

20

30

40

50

得られたパワースペクトルは、音声らしさ推定部 3、雑音スペクトル推定部 4、帯域分離部 5 及び雑音抑圧部 9 にそれぞれ出力される。また、得られた位相スペクトルは周波数・時間変換部 10 に出力される。なお、窓掛け処理としては、例えばハニング窓、台形窓等の公知の手法を用いることができる。また、時間・周波数変換部 2 は窓掛け処理を行う際に、必要に応じてゼロ詰め処理も実施する。FFT は周知の手法であるので説明を省略する。

【0017】

音声らしさ推定部 3 は、時間・周波数変換部 2 から入力した入力信号のパワースペクトルを用いて、現フレームの入力信号の態様が“音声らしいかどうか”の度合いとして、例えば音声の可能性が高い場合には大きな値を取り、音声の可能性が低い場合には小さな値を取るような、音声らしさ評価値を算出する。

10

【0018】

音声らしさ評価値の算出方法として、例えば入力信号のパワースペクトルをフーリエ変換することで得られる自己相関係数の最大値、パワースペクトルの総和から得られる入力信号エネルギー、入力信号の全帯域 SN 比（信号対雑音比）、及びパワースペクトルのばらつき具合を表すスペクトルエントロピ等の公知の手法を、それぞれ単独又は組み合わせて用いることが可能である。ここでは、説明の簡略化のため、現フレームの入力信号のパワースペクトルより計算できる、自己相関係数の最大値を単独で用いた場合について示す。自己相関係数 $c(\tau)$ は、下式 (1) のように求めることができる。

【0019】

$$c(\tau) = F[Y(n, k)] \quad (1)$$

20

ここで、 τ はラグ（遅延時間）、 $F[\]$ はフーリエ変換を表す。このフーリエ変換には時間・周波数変換部 2 で用いたのと同様に、例えば 256 点の FFT を用いることができる。上式 (1) による自己相関係数の算出方法については周知の手法であるので、説明を省略する。

【0020】

音声らしさ推定部 3 は続いて、得られた自己相関係数 $c(\tau)$ を $c(0)$ で除算することで 0 ~ 1 の範囲に正規化を行って、例えば音声の基本周波数が存在する可能性が高い $16 < \tau < 120$ の範囲で自己相関係数の最大値を探索し、得られた最大値を音声らしさ評価値 VAD として雑音スペクトル推定部 4 へ出力する。

30

【0021】

雑音スペクトル推定部 4 は、入力信号のパワースペクトル $Y(k)$ と音声らしさ評価値 VAD とを用いて、入力信号中に含まれる平均的な雑音スペクトルを推定する。具体的には、雑音スペクトル推定部 4 が、音声らしさ推定部 3 の出力である音声らしさ評価値 VAD を参照し、現フレームの入力信号の態様が雑音の可能性が高い場合（即ち音声の可能性が低い場合）、現フレームの入力信号のパワースペクトル $Y(n, k)$ を用いて、雑音スペクトル推定部 4 が記憶している 1 フレーム前の雑音スペクトル $N(n-1, k)$ を更新し、更新した雑音スペクトルを雑音抑圧量生成部 7 へ出力する。

【0022】

雑音スペクトルの更新は、雑音スペクトル推定部 4 が例えば下式 (2) に従って、音声らしさ評価値 VAD が所定の閾値（例えば 0.2）以下の場合に、入力信号のパワースペクトルを雑音スペクトルに反映することで実施する。音声らしさ評価値 VAD が閾値 0.2 を越える場合には、現フレームの入力信号は音声の可能性が高いと考えられるので、雑音スペクトルの更新はせず、1 フレーム前の雑音スペクトルをそのまま現フレームの雑音スペクトルとして用いる。

40

【0023】

$$\begin{cases} \tilde{N}(n, k) = (1 - \alpha(k)) \cdot N(n-1, k) + \alpha(k) \cdot Y(n, k), & VAD \leq 0.2 \\ \tilde{N}(n, k) = N(n-1, k), & VAD > 0.2 \end{cases} \quad (2)$$

ただし、 $0 \leq k < K$

ここで、 n はフレーム番号、 k はスペクトル番号、 K はFFTポイント数の半分の値、 $N(n-1, k)$ は更新前の雑音スペクトル、 $Y(n, k)$ は雑音の可能性が高いと判断された現フレームの雑音スペクトル、 $\tilde{N}(n, k)$ は更新後の雑音スペクトルである。ここでは電子出願の関係上、上式(2)中の“ \sim ”(チルダ記号)を“ $\tilde{\sim}$ ”と表記するが、以降の説明では更新後の雑音スペクトルのチルダ記号を省略する。また、 $\alpha(k)$ は0~1の値を取る所定の更新速度係数であり、比較的0に近い値に設定すると良い。ただし、周波数が高くなるに従って更新速度係数を大きくした方が良い場合もあるので、雑音の種類等に応じて更新速度係数を適宜調整することも可能である。

【0024】

さらに、雑音スペクトル推定部4は、現フレームの雑音スペクトル $N(n, k)$ を、次の更新処理で用いるために記憶しておく。記憶手段としては、例えば半導体メモリ、ハードディスク等に代表されるような、電氣的又は磁氣的な随時読み出し及び書き込み可能な記憶手段を用いる。

【0025】

帯域分離部5は、入力信号のパワースペクトル $Y(k)$ を、非均一な周波数帯域に分割し、サブバンド毎にグループ分けする。図2に、入力信号のパワースペクトル $Y(k)$ の帯域分割の一例を示す。図2の例では、入力信号のパワースペクトル $Y(k)$ の低域から高域までを19個の非均一な周波数帯域に分割し、それぞれのグループをサブバンドにしている。具体的には、サブバンド番号 $z=10$ の場合、そのサブバンドには $k=35 \sim 40$ 番目のスペクトル成分が属することとなる。なお、図2のサブバンドは臨界帯域と呼ばれ、人間の聴覚特性と整合性が高い。この臨界帯域のサブバンド番号の単位はBark(バーク)である。臨界帯域の詳細に関しては、E.ツヴィガー著「心理音響学」(西村書店、1992年8月)を参照することができる。

【0026】

なお、図2の例では臨界帯域で帯域分割する例を示したが、これに限定されるものではなく、例えば低域になるに従って2のべき乗で帯域が狭くなるオクターブバンド分割でも良いし、全ての帯域を例えば4つのスペクトル成分で構成するサブバンドに分割するような均等分割でも良い。また、特定の周波数帯域(低域、音声の重要部分である基本周波数帯域、又はフォルマント成分が分布する可能性が高い帯域)の精度を向上させるために、より細かい単位で分割しても良く、細かい単位で分割することによって後述する雑音抑圧特性の劣化を抑制することができる。帯域分離部5は、以上のように分割処理を実施の後、グループ分けしたサブバンド番号 z 毎のパワースペクトル $Y(z, k)$ を、帯域代表成分生成部6へ出力する。

【0027】

帯域代表成分生成部6は、帯域分離部5から入力したサブバンド毎のパワースペクトル $Y(z, k)$ を用いて、各サブバンドを代表する代表パワースペクトル $Y_d(z)$ を生成して雑音抑圧量生成部7へ出力する。代表パワースペクトル $Y_d(z)$ の生成方法として、例えば下式(3)のように、各サブバンド内でパワースペクトル $Y(k)$ の大きさを順次比較して、最も値が大きなパワースペクトル $Y(k)$ を代表パワースペクトル $Y_d(z)$ とする。ただし、音声らしさ推定部3が出力する音声らしさ評価値 VAD が所定の閾値(例えば、0.2)以下の場合には、代表パワースペクトル $Y_d(z)$ として最も値が大きなパワースペクトル $Y(k)$ を選択する方法ではなく、例えば特許文献1のような、サ

10

20

30

40

50

サブバンド内の全パワースペクトル $Y(k)$ の平均値を算出して代表パワースペクトル $Y_d(z)$ にする方法に切り替える。

【0028】

$$Y_d(z) = \begin{cases} \max_{k=f_1(z)}^{k=f_2(z)} [Y(z, k)] & , \quad VAD > 0.2 \\ \sum_{k=f_1(z)}^{f_2(z)} \frac{Y(z, k)}{(f_2(z) - f_1(z) + 1)} & , \quad VAD \leq 0.2 \end{cases} \quad (3)$$

ただし、 $z = 0, \dots, 18$

10

【0029】

図3は、本実施の形態1の帯域代表成分生成部6の処理効果を模式的に表した図である。図3(a)は、雑音が混入した入力信号の、ある時点でのパワースペクトルをプロットしたグラフであり、縦軸にパワースペクトルの大きさ(振幅)、横軸に周波数を示す。また、実線は入力信号のパワースペクトル成分、破線は雑音スペクトルの包絡線、一点鎖線はサブバンドの境界をそれぞれ表している。さらに、図を簡単にするために、サブバンドは周波数帯域を均等分割した例を示す。

【0030】

図3(b)は、図3(a)に示す入力信号から、従来の方法により各サブバンド内のパワースペクトルの平均値を求め、代表パワースペクトルにした場合の結果を示す。この方法では、音声成分と推測されるパワースペクトルの大きさが小さくなるため、後述する雑音抑圧量生成部7において音声成分が過小評価されることとなり、その結果、音声信号が抑圧されて音声の隠滅感が増大し、音声劣化してしまう。

20

【0031】

一方、図3(c)は、図3(a)に示す入力信号から、帯域代表成分生成部6が代表パワースペクトルを算出した場合の結果を示す。図3の例では入力信号に音声信号が存在しているため、音声らしさ評価値 VAD は閾値 0.2 より十分大きい。このため、帯域代表成分生成部6は式(3)により代表パワースペクトルを求める。図3(c)より、図3(b)の従来の方法に比べて、音声成分と推測されるパワースペクトルは保存され、後段の雑音抑圧量生成部7で音声成分が過小評価されることがなく、音声信号が抑圧されることもない。よって、高品質な雑音抑圧が可能となる。

30

なお、図3ではサブバンドを均等分割した場合について例示したが、例えば図2の表のように臨界帯域幅で非均等分割した場合も同様の効果を奏することはいうまでもない。

【0032】

図3では、音声らしさ評価値 VAD が大きく、かつ、入力信号に音声信号が存在する場合について例示したが、この他、例えば音声らしさ評価値 VAD が小さく、現フレームの入力信号が雑音の可能性が高いと考えられる場合には、大きな値を持つパワースペクトルが存在してもそれは雑音の可能性が高いので、従来の平均値による算出方法に切り替えて代表パワースペクトルを生成するようにしてもよい。サブバンド内のパワースペクトルの平均値を求めるようにすることで、雑音の可能性が高い大きな値のパワースペクトルの振幅が小さくなるので、誤った代表パワースペクトルの生成を抑制することができる。

40

【0033】

なお、入力信号に重畳する雑音が小さい場合等、雑音の影響が少ない場合には、帯域代表成分生成部6が音声らしさ評価値 VAD に応じた代表パワースペクトル算出方法の切り替えを行わず、常に最大値をもつパワースペクトルを代表パワースペクトルにする方法を採用してもよい。

【0034】

雑音抑圧量生成部7は、帯域代表成分生成部6から入力した代表パワースペクトル $Y_d(z)$ と、雑音スペクトル推定部4から入力した雑音スペクトル $N(n, k)$ とを用いて、予め用意された所定の演算式に従ってサブバンド毎の雑音抑圧量 $G(z)$ を生成し、帯

50

域多重化部 8 へ出力する。この雑音抑圧量 $G(z)$ の演算式の導出方法は後述する。

【0035】

帯域多重化部 8 は、雑音抑圧量生成部 7 が求めたサブバンド毎の雑音抑圧量 $G(z)$ を、各サブバンドに属するスペクトル毎に多重化し、スペクトル毎の雑音抑圧量 $G(k)$ に展開する。具体的には、同一のサブバンド番号 z に属するスペクトル番号 k の雑音抑圧量 $G(k)$ の値に、そのサブバンド番号 z の雑音抑圧量 $G(z)$ の値をコピーすることで行う。雑音抑圧量生成部 7 は、得られたスペクトル毎の雑音抑圧量 $G(k)$ を雑音抑圧部 9 へ出力する。

【0036】

雑音抑圧部 9 は、時間・周波数変換部 2 から入力した入力信号のパワースペクトル $Y(k)$ と、雑音抑圧量生成部 7 から入力したスペクトル毎の雑音抑圧量 $G(k)$ とを用いて、下式(4)により、雑音抑圧された入力信号のパワースペクトル $\hat{Y}(k)$ を生成し、周波数・時間変換部 10 へ出力する。電子出願の関係上、上式(4)中の“^” (ハット記号)を“^”と表記し、これ以降に示す式の説明でも“^”と表記する。

【0037】

$$\hat{Y}(k) = G(k) \cdot Y(k) \quad (4)$$

ただし、 $k = 0, \dots, K$

ここで、 K は FFT ポイント数の半分の値である。

【0038】

周波数・時間変換部 10 は、雑音抑圧部 9 から入力した雑音抑圧された入力信号のパワースペクトル $\hat{Y}(k)$ と、時間・周波数変換部 2 から入力した位相スペクトル $P(k)$ とを用いて、逆高速フーリエ変換(逆 FFT)により周波数領域のスペクトルから時間領域の信号に変換し、周波数・時間変換部 10 の内部に記憶している前フレームの信号とオーバーラップ処理した後、雑音抑圧された入力信号 $y^{\wedge}(t)$ として、出力端子 11 に出力する。出力端子 11 はこの雑音抑圧された入力信号 $y^{\wedge}(t)$ を出力する。

【0039】

続いて、雑音抑圧量生成部 7 の演算方法を、図 4 を用いて説明する。図 4 に示す雑音抑圧量生成部 7 は、事後 SNR (信号対雑音比) 推定部 71、事前 SNR 推定部 72、雑音抑圧量計算部 73、及び遅延部 74 を備える。以下、T. Lotter, P. Vary, “Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model” (EURASIP Journal on Applied Signal Processing, Vol. 2005, No. 7, pp. 1110 - 1126, July 2005) に記載されている演算方法(最大事後確率法: Maximum A Posteriori; MAP 法)に基づいて、雑音抑圧量の演算方法を説明する。

【0040】

事後 SNR 推定部 71 は、帯域代表成分生成部 6 から入力した代表パワースペクトル $Y_d(z)$ と、雑音スペクトル推定部 4 から入力した雑音スペクトル $N(k)$ とを用いて、下式(5)によりサブバンド毎の事後 SNR (a posteriori SNR) $\hat{\gamma}(n, z)$ を推定する。ただし、雑音スペクトル $N(z)$ は、サブバンドに対応付けるために、例えば下式(6)に従って求めたサブバンド毎の平均値とする。

【0041】

$$\hat{\gamma}(n, z) = \frac{Y_d(n, z)}{N(n, z)} \quad (5)$$

ただし、 $z = 0, \dots, 18$

10

20

30

40

$$N(z) = \sum_{k=f_1(z)}^{f_2(z)} \frac{N(z,k)}{(f_2(z) - f_1(z) + 1)} \quad (6)$$

ただし、 $z = 0, \dots, 18$

【0042】

事前SNR推定部72は、事後SNR推定部71から入力したサブバンド毎の事後SNR $\hat{r}(n, z)$ と、後述する遅延部74を通じて得られる前フレームの雑音抑圧量 $G(n-1, z)$ とを用いて、下式(7)により事前SNR (a priori SNR) $\hat{\xi}(n, z)$ を再帰的に推定する。なお、事前SNR推定部72は、内部メモリ等の記憶手段に前フレームの事後SNR $\hat{r}(n-1, z)$ を記憶しておき、現フレームでの計算に使用する。

10

【0043】

$$\hat{\xi}(n, z) = \alpha \cdot \hat{\gamma}(n-1, z) \cdot G^2(n-1, z) + (1-\alpha) \cdot F[\hat{\gamma}(n, z) - 1] \quad (7)$$

ただし、 $F[x] = \begin{cases} x, & x > 0 \\ 0, & \text{else} \end{cases}$

20

ここで、 α は $0 < \alpha < 1$ の値を持つ所定の忘却係数であり、 $\alpha = 0.98$ が好適な値として選択可能であるが、入力される音声及び雑音の態様に応じて適宜調整してもよい。

【0044】

雑音抑圧量計算部73は、事前SNR推定部72から入力した事前SNR $\hat{\xi}(n, z)$ と、事後SNR推定部71から入力した事後SNR $\hat{r}(n, z)$ を用いて、下式(8)によりサブバンド毎の雑音抑圧量 $G(z, n)$ を計算し、帯域多重化部8へ出力すると共に、遅延部74へも出力する。

【0045】

$$G(n, z) = u + \sqrt{u^2 + \frac{v}{2\hat{\gamma}(n, z)}} \quad (8)$$

30

ただし、 $u = \frac{1}{2} - \frac{\mu}{4\sqrt{\hat{\gamma}(n, z) \cdot \hat{\xi}(n, z)}}$

ここで、 v 及び μ は所定の係数であり、上述した最大事後確率法に関する文献では $v = 0.126$ 、 $\mu = 1.74$ が好適な値として例示がある。もちろん、この値以外であってもよく、入力信号及び雑音の態様に応じて適宜調整することができる。

【0046】

40

遅延部74は、後述の雑音抑圧量計算部73が出力する、前フレームのサブバンド毎の雑音抑圧量 $G(n-1, z)$ を内部に保持しておき、上式(7)の現フレームの計算に適用するように事前SNR推定部72に送出する。

【0047】

以上より、この実施の形態1によれば、雑音抑圧装置は、入力端子1から入力した時間領域の入力信号を、周波数領域の信号であるパワースペクトルと位相スペクトルに変換する時間・周波数変換部2と、入力信号に重畳した雑音スペクトルを推定する雑音スペクトル推定部4と、時間・周波数変換部2が変換した複数のパワースペクトルをサブバンドにまとめる帯域分離部5と、サブバンド内の複数のパワースペクトルのうち最大値を持つパワースペクトルを代表パワースペクトルにする帯域代表成分生成部6と、代表パワースペ

50

クトルと雑音スペクトルとを用いてサブバンドの雑音抑圧量を算出する雑音抑圧量生成部 7 と、サブバンド毎の雑音抑圧量をスペクトル毎に変換する帯域多重化部 8 と、スペクトル毎に、雑音抑圧量に応じてパワースペクトルの振幅を抑圧する雑音抑圧部 9 と、位相スペクトルと雑音抑圧部 9 で振幅抑圧されたパワースペクトルとを時間領域の信号に変換して出力端子 11 から出力する周波数・時間変換部 10 とを備える構成にした。このため、代表パワースペクトルを用いて雑音抑圧量を算出するので処理量を低減することができる。また、この代表パワースペクトルにはグループ内の値の大きいパワースペクトルを用いるので、雑音抑圧量算出時に入力信号の音声成分が過小評価されることがなくなり、その結果、音声信号を抑圧せず、高品質な雑音抑圧を行うことができる。

【0048】

また、この実施の形態 1 によれば、雑音抑圧装置は、入力信号が音声らしいかどうかの度合いを示す音声らしさ評価値を算出する音声らしさ推定部 3 を備え、帯域代表成分生成部 6 は、音声らしさ評価値に基づいて、入力信号の音声らしさの度合いが高い場合にはサブバンド内の最大値をもつパワースペクトルを代表パワースペクトルにし、当該入力信号の音声らしさの度合いが低い場合にはサブバンド内の複数のパワースペクトルの平均値を求めて代表パワースペクトルを生成するように構成した。このため、誤った代表パワースペクトルの生成を抑制することができ、高品質な雑音抑圧が可能となる。

【0049】

なお、上記実施の形態 1 では、事後 SNR 推定部 71 において、雑音スペクトルをサブバンド毎に対応付けるために式 (6) により平均値を求める構成にしたが、これに限定されるものではなく、例えば、代表パワースペクトル $Y_d(z)$ を生成する際に選択した、最も値の大きいパワースペクトル $Y(k)$ のスペクトル番号 k に対応する雑音スペクトル $N(k)$ を対応付ける構成にしてもよい。この構成の場合、特に帯域分割幅が狭い場合に事後 SNR の推定精度が向上し、更に高品質な雑音抑圧を行うことができる。

【0050】

また、上記実施の形態 1 では、帯域多重化部 8 において、サブバンド毎の雑音抑圧量 $G(z)$ を、同一のサブバンドに属するスペクトル毎の雑音抑圧量 $G(k)$ にコピーすることにより展開する構成にしたが、これに限定されるものではなく、例えば、隣接するサブバンドの雑音抑圧量 $G(z-1)$ 、 $G(z+1)$ を用いて、下式 (9) のように重み付き平均を求めても良い。

【0051】

$$G(z, k) \Big|_{k=f_1(z)}^{f_2(z)} = \frac{(L-k) \cdot G(z-1)}{4L} + \frac{G(z)}{2} + \frac{k \cdot G(z+1)}{4L} \quad (9)$$

この式 (9) により求まる左辺の値は、サブバンド番号 z に属するスペクトル毎の雑音抑圧量 $G(k)$ を意味し、スペクトル番号 k が図 2 の表中の $f_1(z)$ から $f_2(z)$ まで変化することを示す。また、右辺は、サブバンド番号 z の成分に 0.5 の重み付けを行い、隣接するサブバンド番号 $z-1$ 、 $z+1$ の成分にそれぞれ 0.25 の重み付けを行うことを意味し、さらに、スペクトル番号 k の $f_1(z)$ から $f_2(z)$ までの変化に対応して重みが連続的に変化することを表す。L は、サブバンド番号 z に属するスペクトル番号 k の個数を表す。このように重み付き平均をとることにより、特に、帯域分割幅が広い場合に雑音抑圧量 $G(k)$ の周波数方向の変化が安定し、更に高品質な雑音抑圧を行うことができる。

【0052】

また、上記実施の形態 1 では、帯域代表成分生成部 6 が代表パワースペクトルを生成する際に値が最も大きいパワースペクトルを選択しているが、これに限定されるものではなく、例えば、サブバンドの境界付近に値が最も大きいパワースペクトルが存在していたとして、サブバンドの中央付近の周波数に属し、かつ、2 番目に値が大きいパワースペクトルを優先して選択したり、あるいは、上式 (3) を用いたパワースペクトル探索の際に所定の閾値を越えたパワースペクトルを検出した時点で探索を終了して代表パワースペクト

10

20

30

40

50

ルにしたりすることも可能である。

サブバンド中央付近の周波数に属するパワースペクトルを優先して選択することにより、帯域分割幅が広い場合に事後SNRの推定精度が向上する効果がある。また、所定の閾値を越えたパワースペクトルが検出できた時点で探索を終了することで、代表パワースペクトル探索に要する処理量を削減できる効果がある。

【0053】

また、本実施の形態1の音声らしさ推定部3では、音声らしさ評価値として入力信号の自己相関係数の最大値を用いる構成としたが、これに限定されるものではなく、例えば、上述したスペクトルエントロピ等の公知の手法の他、時間領域の入力信号を分析した結果である、線形予測残差パワー等を組み合わせて用いる構成にしてもよい。

10

【0054】

実施の形態2 .

上記実施の形態1では、帯域代表成分生成部6において、同一サブバンド内で最も値が大きいパワースペクトルを代表パワースペクトルに選択していたが、例えば、同一サブバンド内でパワースペクトルを値が大きい順に並び替えて、値が大きなパワースペクトルから大きな重みを付けて重み付き平均を求め、その値を代表パワースペクトルにしてもよい。

また、例えば、メジアン等の統計的手法を用いて、中央値を代表パワースペクトルにしてもよい。

【0055】

20

以上より、この実施の形態2によれば、帯域代表成分生成部6が、サブバンド内の複数のパワースペクトルのうち、値の大きいパワースペクトルから順に大きい重みを付けて求めた重み付き平均を、代表パワースペクトルにする構成にした。このため、高騒音時において音声らしさ評価値の分析精度が低下したり、音声成分と雑音成分の見分けが困難な時に、安定して代表パワースペクトルの生成を行うことができるようになり、高品質な雑音抑圧を行うことができる。

また、重み付き平均に代えて、メジアン等の統計的手法を用いても同様な効果を得られる。

【0056】

実施の形態3 .

30

上記実施の形態1では、帯域代表成分生成部6において、音声らしさ評価値が閾値を越えると同一サブバンド内の最大値を持つパワースペクトルを代表パワースペクトルに選択し、一方、閾値未満なら同一サブバンド内の各パワースペクトルから平均値を求めて、この平均値をもつ代表パワースペクトルを生成するように切り替え制御を行う構成にしたが、例えば、下式(10)のように、音声らしさ評価値VADを重み付け係数にして、最大値と平均値の重み付き和を代表パワースペクトルにすることも可能である。

【0057】

$$Y_d(z) = VAD \cdot \left\{ \max \left[Y(z, k) \right]_{k=f_1(z)}^{k=f_2(z)} \right\} + (1 - VAD) \cdot \left\{ \sum_{k=f_1(z)}^{f_2(z)} \frac{Y(z, k)}{(f_2(z) - f_1(z) + 1)} \right\} \quad (10)$$

40

ただし、 $0 \leq VAD \leq 1$

この式(10)は、音声らしさ評価値VADに応じて、連続的に最大値と平均値を切り替えることが可能である。入力信号が音声の可能性が高い場合には、音声らしさ評価値VADが大きくなるので、代表パワースペクトルは最大値の場合の重みが大きくなる。一方、雑音の可能性が高い場合には、音声らしさ評価値VADが小さくなるので、平均値の場合の重みが大きくなる。

50

【 0 0 5 8 】

以上のように、この実施の形態 3 によれば、帯域代表成分生成部 6 は、音声らしさ評価値を重み付け係数に用いて、サブバンド内の複数のパワースペクトルの最大値と平均値の重み付き和を算出し、代表パワースペクトルにする構成とした。このため、音声成分と雑音成分の見分けが困難な時でも安定して代表パワースペクトルの生成を行うことができるようになり、高品質な雑音抑圧を行うことができる。

【 0 0 5 9 】

実施の形態 4 .

上記実施の形態 1 では、帯域代表成分生成部 6 において、音声らしさ評価値に基づいて全サブバンドの代表パワースペクトル生成の切り替え制御を行っていたが、サブバンド毎に切り替え制御を行っても良い。例えば、帯域代表成分生成部 6 がサブバンド内のパワースペクトルの分散を計算し、分散が所定の閾値を越える場合には、そのサブバンドは音声成分を含むと判断して、代表パワースペクトルとして最大値を選択する方法に切り替える。一方、分散が所定の閾値を下回る場合には、代表パワースペクトルとして平均値を計算する方法に切り替える。

【 0 0 6 0 】

なお、分散は、サブバンド内のパワースペクトルの値のばらつき具合を検出するための 1 方法であって、分散以外にもばらつき具合を検出できる方法であれば別の分析方法を用いても良い。

【 0 0 6 1 】

以上より、この実施の形態 4 によれば、帯域代表成分生成部 6 が、サブバンド毎に代表パワースペクトルの生成方法を切り替えるように構成したので、代表パワースペクトルの生成精度を更に向上することができるようになり、更に高品質な雑音抑圧を行うことができる。

【 0 0 6 2 】

以上の全ての実施の形態 1 ~ 4 では、雑音抑圧量生成部 7 による雑音抑圧の方法として最大事後確率法 (MAP 法) を用いたが、この方法に限定されるものではなく、その他の方法を雑音抑圧量生成部 7 に適用することができる。例えば、非特許文献 1 に詳述されている最小平均 2 乗誤差短時間スペクトル振幅法、又は S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction" (IEEE Trans. on ASSP, Vol. 27, No. 2, pp. 113 - 120, Apr. 1979) に詳述されているスペクトル減算法等がある。

【 0 0 6 3 】

また、以上の実施の形態 1 ~ 4 では、図 2 に示すように、帯域分離部 5 による帯域分割の例として狭帯域電話 (0 ~ 4000 Hz) の場合について説明しているが、雑音抑圧装置の雑音抑圧対象は狭帯域電話音声に限定されるものではなく、例えば 0 ~ 8000 Hz 等の広帯域電話音声又は音響信号でもよい。

【 0 0 6 4 】

また、上記実施の形態 1 ~ 4 において、雑音抑圧された入力信号 $y^{\wedge}(t)$ は、デジタルデータ形式で音声符号化装置、音声認識装置、音声蓄積装置、ハンズフリー通話装置等の各種音響処理装置へ送出されるが、実施の形態 1 ~ 4 の雑音抑圧装置は、単独又は上述の他の装置と共に DSP (デジタル信号処理プロセッサ) によって実現したり、ソフトウェアプログラムとして実行したりすることでも実現可能である。プログラムはソフトウェアプログラムを実行するコンピュータの記憶装置に記憶していても良いし、CD-ROM 等の記憶媒体にて配布される形式でも良い。また、ネットワークを通じてプログラムを提供することも可能である。また、雑音抑圧された入力信号 $y^{\wedge}(t)$ を出力端子 11 の後段で D/A (デジタル・アナログ) 変換し、増幅装置にて増幅し、スピーカ等から直接音声信号として出力することも可能である。

【 産業上の利用可能性 】

10

20

30

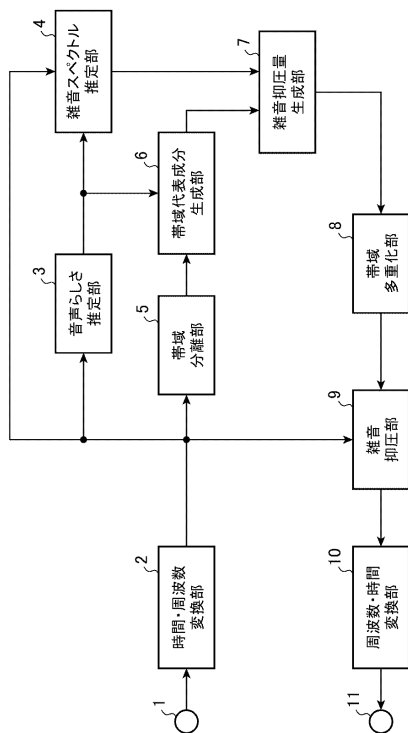
40

50

【 0 0 6 5 】

以上のように、この発明に係る雑音抑圧装置は、少ない処理量で高品質な雑音抑圧を行うようにしたので、音声通信・音声蓄積・音声認識システムが導入された、カーナビゲーション・携帯電話・インターフォン等の音声通信システム・ハンズフリー通話システム・TV会議システム・監視システム等の音質改善、及び、音声認識システムの認識率の向上のために供するのに適している。

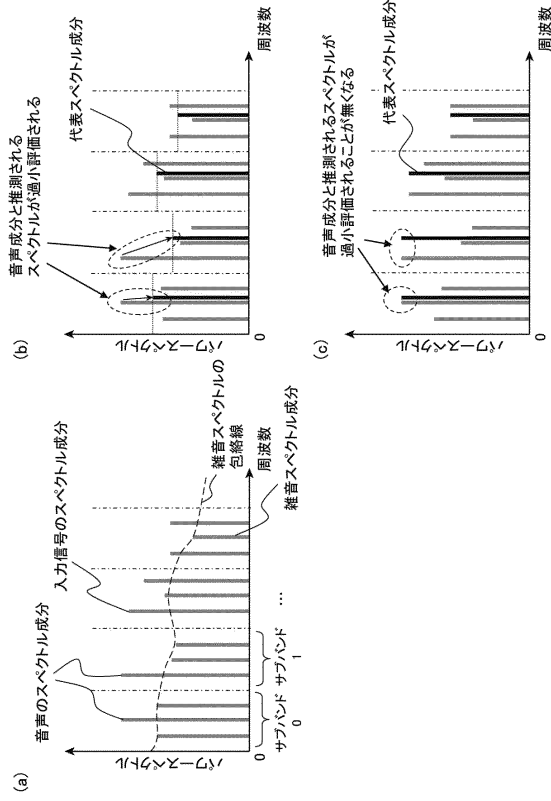
【 図 1 】



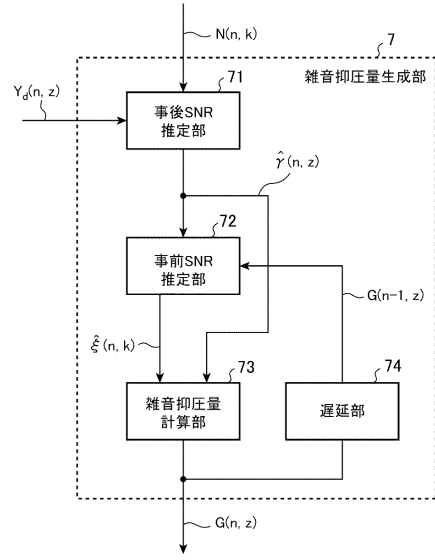
【 図 2 】

サブバンド番号 z	スペクトル番号k		周波数帯域(Hz)
	$f_1(z)$	$f_2(z)$	
0	0	0	0~30
1	1	3	30~90
2	4	6	90~180
3	7	9	180~280
4	10	12	280~375
5	13	16	375~500
6	17	20	500~625
7	21	24	625~750
8	25	29	750~900
9	30	34	900~1050
10	35	40	1050~1250
11	41	47	1250~1470
12	48	55	1470~1720
13	56	64	1720~2000
14	65	74	2000~2310
15	75	86	2310~2680
16	87	100	2680~3120
17	101	118	3120~3690
18	119	127	3690~4000

【図3】



【図4】



フロントページの続き

(72)発明者 田崎 裕久
東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

審査官 千本 潤介

(56)参考文献 国際公開第2009/038136(WO, A1)
特開2010-055024(JP, A)
特開2002-140100(JP, A)
特開2007-183306(JP, A)
特開2006-178333(JP, A)

(58)調査した分野(Int.Cl., DB名)
G10L 21/0232
G10L 21/0264